Available online at www.pelagiaresearchlibrary.com

Pelagia Research Library

European Journal of Experimental Biology, 2014, 4(3):566-575

ISSN: 2248 –9215
CODEN (USA): EJEBAU

Bayesian estimation of heritability and genetic gain for subsets of genotypes evaluated in a larger set of genotypes in a block design

Siraj Osman Omer[1], Abdelwahab Hassan Abdalla[2], Salvatore Ceccarelli[3,4], Stefania Grando[3,4] and Murari Singh[3]

[1]*Experimental Design and Analysis Unit, Agricultural Research Corporation (ARC), Wad Medani, Sudan*
[2]*Department of Agronomy, Faculty of Agriculture, University of Khartoum, Sudan*
[3]*International Center for Agricultural Research in the Dry Areas (ICARDA), Amman, J ordan*
[4]*International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Andhra Pradesh, India*

ABSTRACT

There are situations in which a number of inbred lines are found grouped into classes, depending on their origin and phenology. Interest in such situations lies in the estimation of genotypic variations between the genotypes from individual groups, whereas all genotypes are evaluated in a single trial conducted in a randomized complete block design (RCBD) or an incomplete block design (IBD). The objective of this study was to apply a Bayesian approach to estimate genotypic variation, heritability, and genetic advances within individual groups of groups of genotypes. A set of 360 barley genotypes were evaluated in a two replicate alpha-design with blocks of 10 plots each. The standard frequentist method to estimate variance components was carried out by the restricted maximum likelihood method on the days to flower data from the IBD as well as by an RCBD by ignoring the incomplete blocks. The Bayesian approach with selection of best priors was implemented for the estimation. We noticed a substantial difference in the estimates of the various genetic parameters across the groups. The estimation of variations between the genotypes from individual groups (RCBD or IBD) is needed as the basis of many agricultural research or plant breeding/agronomy trials. The Bayesian approach uses broader inference framework to integrate the prior information on parameters with the likelihood of the current data. Therefore, the Bayesian approach presented here for estimation of heritability and genetic gain for subsets of genotypes evaluated in a larger set of genotypes in the block design is recommended for use in similar situations.

Keywords: Bayesian estimation, genotypic variance, heritability, genetic advance.

INTRODUCTION

In self-pollinated crops, plant breeders develop inbred lines (genotypes) over a course of time[15].In many situations, these genotypes can be traced back to have arisen from a number of distinct crosses[17]. The genotypes may also be grouped according to genetic make-up, e.g., being derived from the same cross, genetic marker class, seed colour, some phonological characteristics such as flowing time - early, medium and late flowering lines, and maturity time – early, medium or late maturing lines, etc.[11].Eventually, the groups may represent different target objectives or target environments as in breeding programs based on decentralized selection [4]. The interest lies in estimating the genetic parameters based on the genotypes within each group, while the evaluation of all the genotypes is done using a single trial in randomized complete block design (RCBD) or incomplete block design IBD[1]. The genotypes of the individual groups are not evaluated in separate trials[10].

Two approaches for estimation of parameters, frequentist and the Bayesian can be pursued.In Bayesian analysis, prior assumptions and the likelihood of the data at hand form the joint posterior density of all unknown variables in a model underlying the observed phenotypes[8]. The prior information is generally available in an ongoing Crop Improvement Program. Markov chain Monte Carlo (MCMC) methods can be used for exploration of complex non-standard joint densities and marginal posterior densities of parameters of interest in genetic gain for subsets of genotypes approximated[5]. There are variety of techniques for their implementation[7]of which Gibbs sampling[9] is the most commonly used in the Bayesian approach analyses and in combination with MCMC methods, have been applied in human genetics[21]and in plant genetics[16]genotyped individuals[12]. Heritability (h^2) is one of the genetic parameters, which measures variability that can be expressed in terms of the phenotypic variability existing in a population of inbred lines and also determines the expected genetic advance due to selection of desired genotypes[6].A Bayesian approach for estimating heritability's using Markov-chain Monte Carlo (MCMC) simulationshas been given in [18].The objective of this paper is to present estimation of genetic parameters such as heritability and genetic advance for each group of genotypes by due partitioning of total genotypic variability into variability within and between the groups, under frequentist and the Bayesian approach. This investigation may shed light on a hypothesis that the magnitude of heritability does not only depend on the environment (it is often said that in stress environment heritability is lower than in non-stress environments, even if this is not necessarily true as in Table 3 of [2,3],but also on the difference in adaptation between the parents. The application was made on barley datasets where genotypes were derived from four crosses[13]. Estimation of heritability and genetic gains for subsets of genotypes evaluated in a larger set of genotypes in block design were evaluated in the same trial [14].Therefore, evaluating groups of genotypes in the same block design is an efficient cost-effective operation and to ensure the best possible genetic material is identified [22].

MATERIALS AND METHODS

EXPERIMENTAL DESIGNS AND DATASETS

A set of 360 barley genotypesderived from four crosses were evaluated in analpha-design with blocks of 10 plots eachand two replications during 1998/99 at Tel Hadya, Aleppo, Syria. From each cross, we grew 88 RIL's and the two parents. Plot-wise data on days to heading were analyzed.The four populations, one from each cross, were chosen among twelve mapping populations developed at ICARDA by advancing through single seed descentusing parents with various combinations of adaptation to Syrian dryland conditions.. Plot size was 8 rows at 20 cm distance and 2.5 m long. We represent the crosses or the groups of genotypes by 3, 6, 7 and 9as given in Table 1.

Two datasets referred here are: incomplete block design [IBD] dataset, Dataset-IBD (consisting of columns of replications, incomplete blocks within replications, crosses/groups of genotypes, days to heading), and randomized complete block design [RCBD] dataset, Dataset-RCBD obtained from Dataset-IBD by removing the column of the incomplete blocks.

Table 1. The four populations of RIL's used in 1998/99

Group	Cross	Name	Combination of *	Row type
1	3	WI2269/Line 251-11-2/3/Leb71/CBB37//Leb71/CBB29	A ×NA	2
2	6	Gustoe/6/M64-76/Bon//Jo/York/3/M5/Galt//As46/4/Hj34-80/ Astrix/5/NK1272	NA × NA	6
3	7	Arta/3/Harmal-02//Esp/1808-4L	A ×A	2
4	9	Zanbaka/5/Pitayo/Cam//Avt/RM1508/3/Pon/4/Mona/Ben//Cam	A ×NA	2

** A= Adapted and NA = not adapted parents*

ESTIMATION OF HERITABILITY AND GAIN DUE TO SELECTION WITHIN GROUPS OF GENOTYPES

Let the total of v genotypes be grouped into s groups, with v_k genotypes in group k, k =1, …, s.Heritability and gain due to selection are expressed in terms of estimates of genotypic variance (σ^2_{gk})within group kand common experimental error (environmental) variance (σ^2_e) obtained from modeling of all the v genotypes under the chosen experimental design, RCBD or IBD[19].Heritability,in broad-sense and on a mean-basis, is given by$h^2_k = \sigma^2_{gk}/(\sigma^2_{gk} + \sigma^2_e/r)$, where $r$is the number of replications and $k = 1,..,s$ where $s$is the number of groups. Assuming normal distribution for the trait, the genetic gain due to selection for the group k, GA (p)%, at selection intensity p is givenby $\%GA_k(p) = 100\, C(\sigma^2_{gk}/\bar{Y})/(\sigma^2_{gk} + \sigma^2_e/r)^{1/2}$, where $0 < p < 1$, $C = \frac{1}{p\sqrt{2\pi}} e^{-z_p^2/2}$. The

truncation point z_p in the standard normal distribution is given by the equation $\int_{z_p}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 1 - p$, $\bar{Y}$ is the trial or location mean. For p = 0.20, C = 1.4 [17].

ESTIMATES OF VARIANCE COMPONENTS FOR GENOTYPEEFFECTS

Let the genotype factor within group k be represented by abbreviated textGeno$_k$ and the group factor by Grp (k=1,2…,s).The replications are denoted by Rep and incomplete blocks within replications by Blk.Rep. The model for variance component was fitted in terms of additive random effects factors 1) Rep, Grp and Geno1, Geno2, …, Genos for data from RCBD, and 2) Rep, Blk.Rep, Grp and Geno1, Geno2, …, Genosfor data from incomplete block design. Furthermore, the effects of the factors in above order and plot-errors were assumed to follow normal distribution with mean zero and variances, $\sigma_r^2, \sigma_{grp}^2, \sigma_{g1}^2, ..., \sigma_{gs}^2$ and σ_e^2 respectively for RCBD and σ_r^2, $\sigma_b^2, \sigma_{grp}^2, \sigma_{g1}^2, ..., \sigma_{gs}^2$ and σ_e^2 respectively for IBD datasets.

BAYESIAN ESTIMATION

The R2WinBUGS software was the environment for MCMC simulation using the Gibbs sampler[20]. The assumed *a priori* distributions for various variance components are listed in the following for the two models.

For the variance components from RCBD:

1)P$_1$: the priors for the standard deviation components $\sigma_r, \sigma_{grp}, \sigma_{g1} \ldots \sigma_{gs}$ and σ_e follow Uniform (0, θ =50)or Uniform (0, θ =100)

2)P$_2$: the priors for the standard deviation components $\sigma_r, \sigma_{grp}, \sigma_{g1} \ldots \sigma_{gs}$ and σ_e follow positive half-normal distribution $N(0, \tau^{-1} = 100)$. Here, τ is precisionparameter, $\tau = \sigma^{-2}$ given as inverse of variance. In the notation of WinBUGS, this distribution is denoted as dnorm(0, τ)I(0,) or dnorm(0, 0.01)I(0,). I(0,) denotes that only the positive values were taken.

3)P$_3$: the priors for the standard deviation components $\sigma_r, \sigma_{grp}, \sigma_{g1} \ldots \sigma_{gs}$ and σ_e follow positive Half-t distribution $dt(0, c, \upsilon) I(0,)$. Here, c is non-centrality parameter and υ is the degree of freedom of the t-distribution. The values of c and υ are set at 5 and 2 respectively.

4)P$_4$: the priors of P$_3$ with c = 2 and υ = 4.

For variance components from IBD:

1) Q1: the priors for the standard deviation components $\sigma_r, \sigma_b, \sigma_{grp}, \sigma_{g1} \ldots \sigma_{gs}$ and σ_e follow Uniform (0, 10) or Uniform (0, 100) or Uniform (10, 100) or Uniform (100, 1000)

2) Q$_2$: the priors for the standard deviation components $\sigma_r, \sigma_b, \sigma_{grp}, \sigma_{g1} \ldots \sigma_{gs}$ and σ_e follow , the positive Half-normal (0.5, 0.01), Half-normal (0.5, 0.5), Half-normal (0.5, 0.1), or Half-normal (0.1, 0.1) .

3) Q$_3$: The priors for the standard deviation components $\sigma_r, \sigma_b, \sigma_{grp}, \sigma_{g1} \ldots \sigma_{gs}$ and σ_e follow Half-t distribution $dt(0, c, \upsilon) I(0,)$ here, c is non-centrality parameter and υ is the degrees of freedom of the t-distribution. The values of c and υ are set at Half-t (0, 2, 3), Half-t (0,5, 2) and Half-t (0,5, 3).

RESULTS

Selection of priors and estimates of Heritability and Genetic Advance for Dataset-RCBD

The standard frequentist analysis of RCBD dataset was carried out by REML for estimating variance components. Selection of the best priors for Bayesian analysis was carried out using the discrepancy statistics given in Table 2. For the priors sets, P$_1$ to P$_2$, (uniform and Half normal distributions), the values of deviance information criterion (DIC) and effective number of parameters were close.However, the prior set P$_2$ seems to have numerically lowest value of DIC (4418.72). We took P$_2$ for further estimation of the heritability parameters.

Table 2. Deviance information criterion (DIC) values for selection of the priors for Dataset-RCBD

Priors set	$\bar{D}$	$\hat{D}$	p_D	DIC
P_1	4259.22	4099.02	160.197	4419.41
P_2	4261.33	4103.95	157.386	4418.72
P_3	4299.05	4166.95	132.099	4431.15

$\bar{D}$=posterior mean of (- 2 × log-likelihood). $\hat{D}$ = - 2 × log-likelihood at posterior means of parameters. p_D = effective number of parameters, DIC = Deviance information criterion. Priors set are:

P_1: σ_r, σ_{grp} , $\sigma_{gl}, \dots, \sigma_{gs}$ and σ_e independently $\sim$ uniform$(0, 10)$

P_2: σ_r, σ_{grp} , $\sigma_{gl}, \dots, \sigma_{gs}$ and σ_e independently $\sim$ half $-$ normal $(0.1, 0.1)$

P_3: σ_r, σ_{grp} , $\sigma_{gl}, \dots, \sigma_{gs}$ and σ_e independently $\sim$ halft$(0, 5, 2)$

Table 3 gives estimates of variancescomponents and coefficients of variation to measure field heterogeneity under frequentist and Bayesian approaches. It also gives estimates of genotypic variance and heritability on mean-basis, genetic gain due to top 20% selection and a test for significance of genotypic variance. Bayesian estimate of heritability (posterior expected value of the variance competent of frequentist approach especiallyvarianceof genotypes (group 1) was zero while Bayesian approach recorded the variance of genotypes group 1 as (1.543), this means that Bayesian approachgave an estimatebased on genotypic variance which is less likely to be fixed at a boundary.. The MC errors for all the parameters in Table 3 were very small indicating reliable numerical approximation based on 50000 iterations, 5000 simulation runs and three chains. It can be seen that distribution of variance components and heritability from Bayesian approach are skewed based on difference between their means and their variances. Their median values for the Bayesian estimateare very close to the frequentist estimates of GA(.20)%for genotypes Group 2, Group3 and Group4 as (2.03 vs.1.950), (3.88 vs.5.256)and (5.18 vs. 5.16)respectively, while GA(.20)% of genotypes (Group 1) was(0.0001 vs.0.625). The median estimates for genetic gain are very close, under frequentist approach and Bayesian approach of genotypes Group 2, Group 3 and Group 4 as (0.536 vs. 0.652), (0.336 vs. 0.352) and(0.635 vs. 0.649) respectively, while the genotypes of Group1 is totally different as (0.0001 vs. 0.135). Based on the genotypic variance, there is statistically significant variation in the genotypic means (P<0.01) under both theapproaches respectively.There wassubstantial variation due to the groups of the genotypes in the Bayesian estimates: 1.543 –18.93 for genotypic variance component, 0.113 -0.627 for heritability and 0.625 - 5.161% for genetic gain. Thus the selection of lines from a desired group or cross matters, e.g., Cross 6 (Group 2) gave highest genetic advance of 5.26% in RCBD.

Table 3. Frequentist and Bayesian estimates of variances components, heritability and genetic gain for days to heading using Dataset-RCBD

Parameters[@]	Frequentist approach		Bayesian approach (priors set : P_2)					
	Estimate	SE	Estimate	SE	MC error	Percentile		
						2.50%	50% (median)	97.50%
h_1^2	0.000	0.000	0.113	0.102	0.006	0.001	0.088	0.369
h_2^2	0.536	0.075	0.633	0.063	0.002	0.502	0.637	0.750
h_3^2	0.336	0.108	0.315	0.114	0.005	0.074	0.322	0.516
h_4^2	0.635	0.059	0.627	0.061	0.002	0.495	0.632	0.731
$GA(.2)_1$%	0.00	---	0.625	0.612	0.036	0.005	0.459	2.219
$GA(.2)_2$%	3.88	---	5.256	0.896	0.032	3.641	5.204	7.154
$GA(.2)_3$%	2.03	---	1.950	0.820	0.034	0.381	1.944	3.551
$GA(.2)_4$%	5.18	---	5.161	0.848	0.027	3.564	5.158	6.919
σ_e^2	21.52	1.44	21.79	1.63	0.06	18.91	21.65	25.19
σ_{g1}^2	0.001	0.001	1.543	1.638	0.091	0.012	1.067	5.901
	12.44	3.52	19.470	4.818	0.174	11.540	18.970	30.450
	5.44	2.51	5.373	2.611	0.102	0.887	5.169	10.870
	18.73	4.42	18.930	4.473	0.144	11.150	18.630	29.380
σ_r^2	0.55	0.86	4.576	8.125	0.240	0.065	1.389	28.210
Mean	93.29	2.380	92.840	0.249	0.008	92.350	92.840	93.310

[@] *h_k^2= heritability from group k (k= 1…4), SE= standard error, MC error: Monte Carlo error, $GA(.2)_k$ %: genetic advance at 20% selection intensity for group k, k=1…4.*

Selection of priors and estimates of Heritability and Genetic Advance for Dataset-IBD

The frequents analysis components of Dataset-IBD was carried out byREML variance components usingGenStatand WinBUGS program codes along with the statistics using Bayesian approach. The choices of priors for Bayesian analysis were made fromthe statistics given in Table 4. The values of DIC and pD varied over the three priors sets,4405.03 (for Q_1) 4388.54(for Q_2) and 4387.02 (Q_3).Since the prior set Q_3gavethe lowest value of DIC, we took $Q_3$for further estimation of the genetic parameters.

Table 4. Deviance information criterion (DIC) values for selection of the priors variance components for days to heading using Dataset-IBD

Priors set	$\bar{D}$	$\hat{D}$	p_D	DIC
Q_1	4268.27	4131.5	136.765	4405.03
Q_2	4217.43	4046.32	171.11	4388.54
Q_3	4217.84	4048.66	169.179	4387.02

$\bar{D}$=posterior mean of (- 2 × log-likelihood). $\hat{D}$ = - 2 × log-likelihood at posterior means of parameters. p_D = effective number of parameters, DIC = Deviance information criterion. Priors set are:

Q_1: $\sigma_r, \sigma_b, \sigma_{grp}$, $\sigma_{gl}, \dots, \sigma_{gs}$ and σ_e independently $\sim$ uniform$(0, 100)$

Q_2: $\sigma_r, \sigma_b, \sigma_{grp}$, $\sigma_{gl}, \dots, \sigma_{gs}$ and σ_e independently $\sim$ half $-$ normal $(0, 0.01)$

Q_3: $\sigma_r, \sigma_b, \sigma_{grp}$, $\sigma_{gl}, \dots, \sigma_{gs}$ and σ_e independently $\sim$ half $-$ t$(0, 5, 2)$

Table 5 shows the posterior means of frequentist and Bayesian estimates of heritability and genetic gain for Dataset-IBD. The Bayesian estimate of genetic variance based on mean value is slightly higher (9%) than that under the Frequentist approach with highly heritability and gain selection parameters. Their median values for the Bayesian estimateare very close to the frequentist estimates, of GA (.20)% of genotypes Group 2, Group 3 and Group 4 as (3.891, frequentist vs. 3.903, Bayesian), (2.039 vs. 2.028) and (5.185 vs. 5.367)respectively, while GA(.20)% for Group1) as (0.075 vs. 0.350), is different comparing with others, in another words the GA(.20)% of Group 2 was the best group of genotypes comparing to other groups. The median estimates for heritability are very close to each other, under frequentist and Bayesian approach of genotypes for groupsas (0.016 vs. 0.046), (0.549 vs.0.540), (0.346 vs. 0.332) and (0.647 vs. 0.650)respectively, the genotypic estimatesof heritability of genotypes groupsare heterogeneity also Group4 was given highest value in comparing to others. The Bayesian approach based on priors under model $Q_3$was statistically significant at P<0.05 under the two approaches, while the significance was relatively greater in Bayesian approach. There were substantial variation in the estimates of heritability and genetic gain across the crosses/groups of the genotypes: range 0.046 -0.650 for heritability and 0.235 -5.367% for genetic advance. Furthermore comparing the results of Tables 3 and 5, we notic clearly the importance of experimenting with incomplete blocks.

Table 5. Frequentist and Bayesian estimates of variances components, heritability and genetic gain for days to headings using Dataset-IBD

Parameters[@]	Frequentist approach		Bayesian approach (priors model : Q_3)					
	Estimate	SE	Estimate	SE	MCMC error	percentile		
						2.50%	50% (median)	97.50%
h_1^2	0.016	0.173	0.046	0.065	0.004	0.000	0.015	0.229
h_2^2	0.549	0.077	0.540	0.079	0.002	0.373	0.548	0.674
h_3^2	0.346	0.112	0.332	0.124	0.006	0.053	0.347	0.544
h_4^2	0.647	0.060	0.650	0.060	0.002	0.521	0.655	0.752
$GA(.2)_1$%	0.075	---	0.235	0.350	0.022	0.000	0.074	1.263
$GA(.2)_2$%	3.891	---	3.903	0.835	0.025	2.326	3.894	5.599
$GA(.2)_3$%	2.039	---	2.028	0.884	0.038	0.264	2.044	3.745
$GA(.2)_4$%	5.185	---	5.367	0.880	0.026	3.721	5.340	7.113
Trial mean	93.28	2.6	93.21	0.168	0.006	92.890	93.210	93.530

[@] *h_k^2 = heritability from group k (k= 1…4), SE= standard error, MC error: Monte Carlo error, $GA(.2)_k$ %: genetic advance at 20% selection intensity for group k, k=1…4.*

DISCUSSION

In crop improvement programs, normally a number of inbred lines or genotypes are evaluated in block designs and genetic gain is evaluated for a chosen intensity of selections. Parameters such as genotypic variance and heritability are evaluated for the situation where genotypes are assumed to have arisen from a common genetic pool. In breeding programs, often the several different types of crosses are made, say between parents adapted or non-adapted to different stresses, and the lines are developed by selections and crossing or selfing over the generations. It is of paramount interest to study the performance of lines and developing selection strategies with a cross type and estimate the genetic parameters within each group. The group of genetic material may also be represented by phenology, days to flower and maturity. However, in practice one may not evaluate the genetic materials from different crosses in separate experiments. This study addresses mainly three features of genetic evaluation— 1) parameters of interest vary with the group, 2) all the genotypes arising from various groups are evaluated in randomized complete block or incomplete block design, and 3) availability of prior information on variance components are available from the long-term ongoing experimentation.

The study demonstrated a statistical analysis procedure addressing the complex integration of the above three aspects. The parameters such as genetic variance, heritability and genetic advance were estimated under Bayesian approach for each of the two commonly used experimental designs: an incomplete block design or a large number of genotypes, and a complete block design for relatively a moderate number of genotypes. The priors for standard deviation components were screened from the class of recommended priors (Gelman 2006). In the two situations presented here, priors based on uniform and positive Half-t distributions were found most suited prior out of the three priors considered. The resulting posteriors of these components could be used as the priors in later study to make the analysis more informative. The differences were obvious between frequentist and Bayesian approaches. One may choose Bayesian approach as it integrates the prior information with current data while the frequentist approach does not. It has also been differences in heritability and genetic advance estimates (Tables 3 and 5) arising due to accounting of the incomplete blocks, particularly for Cross 6 (Group 2), implying the need for experimenting in small size blocks for a better accountability of local variation. Extension of the method would be worth exploring for other variants of experimental design and data model.

CONCLUSION

The Bayesian approach integrates the prior information with likelihood of the current data. It has been applied to estimate genetic parameters based on the genotypes from individual groups, while the genotypes from all the groups were evaluated in the same single trial conducted in a randomized complete or incomplete block design. Comparison with frequentist approach made. It demonstrates a step-wise procedure from selection of the best priors and estimation of heritability and genetic advance with help of codes in WinBUGS and R- packages. The procedure presented here is recommended for use in a similar situation in genetic studies.

Acknowledgements
First author is grateful to ICARDA and Arab Fund for Economic and Social Development (AFESD) for granting a fellowship for carrying out the research study.

REFERENCES

[1] Baum M, Grando S, Backes G, Jahoor A, Sabbagh A, Ceccarelli S, *Theoretical and Applied Genetics*, **2003**, 107,7,1215–1225.
[2] Ceccarelli, S, *Euphytica*,**1996**, 92: 203-214
[3] Ceccarelli, S. & Grando, S. *Euphytica*,**1991,** 57,157-219.
[4] Ceccarelli, S, Erskine, W, Grando, S. and Hamblin, J., *Experimental agriculture*, **1994**, 30: 177-187
[5]Eaves L, Erkanli A, Markov *Behavior Genetics*, **2003**,33, 279-299.
[6] Farshadfar E, Romena H, Safari H, *International Journal of Agriculture and Crop Sciences*, IJACS, **2013**,5,9,1015-1021.
[7] Gelfand AE, Dey DK, *J Roy Statist Soc B*, **1994**, 56.
[8]Gelman A, *Bayesian Analysis*, 1, **2006**, 515–533
[9] Geman S, Geman, D, *IEEE Trans.FAMI*, **1984**, 6,721-741.
[10]Gilmour AR, Cullis BR, Verbyla AP, *J. Agric. Biol. Env. Statistics*, **1997**, 2, 269-293.
[11]Habtamu S Million F, Multivariate analysis of some Ethiopian field pea (Pisum sativum L.) genotypes, International Journal of Genetics and Molecular Biology, **2013**, 5, 6, 78-87.
[12] Heath S, Markov chain Monte Carlo segregation and linkage analysis for oligenic models, Am J Hum Genet, **1997**, 61,748-760.
[13]Kent ME, Field trial design in plant breeding, **2009**.
http://imbgl.cropsci.illinois.edu/school/presentations/2009, Eskridge.pdf
[14]Malhotra, RS, Singh M, Erskine W, *J. Indian Soc. Agric. Statist*, **2004,** 57, 71-83.
[15] Sarker A, Singh M, Rajaram S. Erskine W, *Crop Sci*. **2010**,50, 1250 –1259.
[16]Satagopan JM, Yandell BS, Newton MA, *Genetics*.**1996**,144: 805-816
[17] Singh M, Tadesse W, Sarker A, Maalouf F, Imtiaz M, CapettiniF, Nachit M, *Crop Science*, **2013**, 53,811-818.
[18]Shepherd RK, Meuwissen THE, Woolliams JA,. *Proc. Assoc. Advmt. Anim. Breed. Genet*, **2009,**18:84-87.
[19]Singh, M, Ceccarelli S, *Theoretical and Applied Genetics*, **1995**, 90(1):142-5.
[20]Sturtz S, Ligges U, Gelman A, *Journal of Statistical Software*, **2005**, 12, 3, 1–16.
[21]Thomas DC, Cortessis V, *Hum. Hered*, **1992**, 42,63–76.
[22]Van Es HM, Gomes CP, Sellmann M, van Es CL, *Geoderma*, **2007**,140, 346–352.

Appendix

A.1: R- codesforreading data from an RCBD dataset and calling the 'bugs' function

```
#load packs
library(lattice)
library(coda)
library(R2WinBUGS)
#data from comb.....................................
Gdata<- read.table("GroupDataRCB.txt", header=TRUE)
Gdata
rp<- Gdata$Rep   # rp for replication vector
grp<-Gdata$Grp # grp for group vector
gn<- Gdata$Geno  # gn for genotype vector
y<-Gdata$dh
print(cbind(rp, gn,grp, y))
NR<- 2
NG<- 360
NGRP<- 4
NR1<- NR-1
N<- NR*NG          # number of observations
print(cbind(NR,NG, N,NGRP))
mn<- mean(y)
mn
# define the first and last genotype numbers in the groups
first<- c(1, 91,  181, 271)
last<- c(90, 180, 270, 360)
#---------------------------------
data<- list("mn", "y","rp","gn","N", "NR","NGRP", "first","last")
data
inits1<-  list(m=2, rho=c(rep(.01,NR)), g=c(rep(.01,NG)),grp=c(rep(.1,NGRP)),  sig.e=1,  sig.r=1.0, sig.grp=1.0, sig.g=c(rep(.5,NGRP)) )
inits2<-  list(m=1, rho=c(rep(.01,NR)), g=c(rep(.01,NG)),grp=c(rep(.11,NGRP)),   sig.e=1.2,  sig.r=1.0,sig.grp=1.1,  sig.g=c(rep(.5,NGRP)) )
inits3<-  list(m=1, rho=c(rep(.01,NR)), g=c(rep(.01,NG)),grp=c(rep(.12,NGRP)),   sig.e=1.1,  sig.r=1.0,sig.grp=1.0,  sig.g=c(rep(.5,NGRP)) )
inits<- list(inits1, inits2, inits3)
inits
#Step 6
parameters <- c("m", "sig2.r", "sig2.grp", "sig2.g", "sig2.e", "h2", "GA20")
parameters
GroupHeritRCB.sim<- bugs(data, inits, parameters, "GroupHeritRCB.bug", n.chains=3, n.iter=50000, n.sims=5000, debug=TRUE)
#Step 6
```

A.2:WinBUGS codesto model data from RCBD and estimation of heritability and genetic gain

```
# RCBD data analysis
model {
for (i in 1:N){
                              y[i] ~ dnorm(mu[i] , tau.e)
mu[i]<-  m + rho[rp[i]] + g[gn[i]]
                    }
for (i in 1: (NR - 1) ) { rho[i] ~ dnorm(0, tau.r) }

rho[NR]<-   -sum(rho[1:(NR -1)])
# Express genotype effects as group effect and genotype within group effects
# Thus  g() = grp() + grp.g()
for(j in 1: (NGRP - 1) ){
for (i in first[j]: (last[j]-1) ) { g[i] <- grp[j]+grp.g[i]}
g[last[j]]<-  -sum(grp.g[first[j]:(last[j]-1)] ) # the last geno in the group
                    }
   # For the last Group
grp[NGRP]<-  -sum(grp[1:(NGRP-1)])
for(j in NGRP:NGRP){
for (i in first[j]: (last[j]-1) ) { g[i] <- grp[j]+grp.g[i]}
g[last[j]]<-  -sum(grp.g[first[j]:(last[j]-1)] )  # last geno in the last group
}
tau.e<- 1/(sig.e*sig.e)
tau.r<- 1/(sig.r*sig.r)
tau.grp<- 1/(sig.grp*sig.grp)

for(j in 1:NGRP){ tau.g[j] <- 1/(sig.g[j]*sig.g[j]) }
for(j in 1:NGRP){ sig2.g[j] <- (sig.g[j]*sig.g[j])   }
                  sig2.e <- (sig.e*sig.e)
                  sig2.r<- (sig.r*sig.r)
                  sig2.grp<- (sig.grp*sig.grp)
#priors
m ~ dnorm(0.0, 1.0E-6)
for(j in 1:(NGRP-1) ){grp[j]~ dnorm(0, tau.grp)}    # for group effects, except the last grpoup
for(j in 1: NGRP ){
for (i in first[j]: (last[j]-1) ) { grp.g[i] ~ dnorm(0, tau.g[j]) } # geno, except the last geno in the group
                                  }
sig.e ~ dt(0, 5, 2)I(0,)
for(j in 1:NGRP){     sig.g[j] ~ dt(0, 5, 2)I(0,) }
sig.r ~ dt(0, 5, 2)I(0,)
sig.grp~ dt(0,5,2)I(0,)

# Prediction of parameters of interest-- means, heritability, SEs
for(j in 1:NGRP) {  h2[j]<- tau.e/(tau.e + tau.g[j]/NR) }
                # this heritability is on  mean-basis, 8 Jan 2013

    # gain due to 20% selection & 10%, 5% selection K=1.4, 1.755,2.063
for(j in 1:NGRP) { GA20[j] <- 100*1.4/mn/sqrt(tau.g[j]*(1+tau.g[j]/NR/tau.e))  }

      }
```

A.3: R- codes for reading data from an IBD dataset and calling the ‘bugs’ function

```
#load packs
library(lattice)
library(coda)
library(R2WinBUGS)
#data from comb.....................................
#-----------------------------
Gdata<- read.table("GroupDataIBD.txt", header=TRUE)
Gdata
rp<- Gdata$Rep   # rp for replication vector
bl<-Gdata$Blk    # bl for block vector
grp<-Gdata$Grp # grp for group vector
gn<- Gdata$Geno  # gn for genotype vector
y<-Gdata$dh

print(cbind(rp, bl, gn,grp, y))
NR<- 2
NB<- 36
NG<- 360
NGRP<- 4
N<- NR*NG          # number of observations
NBR1<- NR*(NB-1)    # no of generated block effects (NB1 from each repl.: last bloack effect is calculated)

print(cbind(NR,NB,NK,NG, N, NBR1, NGRP))
mn<- mean(y)
mn
#---------------------------------
first<- c(1, 91, 181,271)
last<- c(90, 180, 270, 360)
print(cbind(first, last))
#---------------------------------
data<- list("mn", "y","rp","bl","gn","NR", "NB", "N","NGRP", "first", "last")
data
inits1<-  list(m=2, rho=c(rep(.01,NR)), bet= c(rep(.02, NBR1)), g=c(rep(.01,NG)), grp=c(rep(.1,NGRP)), sig.grp=1.1, sig.e=1,  sig.r=1.0,
sig.b=.53, sig.g=c(rep(.5,NGRP)) )
inits2<-  list(m=2, rho=c(rep(.01, NR)), bet= c(rep(.01, NBR1)), g= c(rep(.02,NG)),grp=c(rep(.1,NGRP)), sig.grp=1.1, sig.e=1.1, sig.r=1.15,
sig.b=.68, sig.g=c(rep(.52,NGRP)) )
inits3<-  list(m=2, rho=c(rep(.02,NR)), bet= c(rep(.02, NBR1)), g=c(rep(.01,NG)),grp=c(rep(.11,NGRP)), sig.grp=1.0,   sig.e=1.05, sig.r=1.25,
sig.b=1.35, sig.g=c(rep(.51,NGRP)) )
inits<- list(inits1, inits2, inits3)
inits
parameters <- c("m", "sig2.grp", "sig2.r", "sig2.g", "sig2.e", "sig2.b","h2", "GA20")

parameters

GroupHerit.sim<- bugs(data, inits, parameters, "GroupHeritIBD.bug", n.chains=3, n.iter=50000, n.sims=5000, debug=TRUE)
#Step 6
```

A.4: WinBUGS codes to model data from IBD and estimation of heritability and genetic gain

```
# IBD data analysis
model {
for (i in 1:N){
                              y[i] ~ dnorm(mu[i] , tau.e)
mu[i]<-  m + rho[rp[i]] + bet[rp[i],bl[i] ] + g[gn[i]]
                    }
  # rho[1...NR-1] Rep effects  result into rho[NR]
for (i in 1:(NR -1)) { rho[i] ~ dnorm(0, tau.r)}
rho[NR]<- - sum(rho[1:(NR-1)])
   for (i in 1:NR) { for(j in 1: (NB-1) ) { bet[i,j]~ dnorm(0,tau.b) }    # except the last, block effects within repl.
bet[i,NB]<- -sum(bet[i,1:(NB-1)])   # last block in the replication
                    }
# Express genotype effects as group effect and genotype within group effects
# Thus  g() = grp() + grp.g()
for(j in 1: (NGRP - 1) ){
for (i in first[j]: (last[j]-1) ) { g[i] <- grp[j]+grp.g[i]}
g[last[j]]<-  -sum(grp.g[first[j]:(last[j]-1)] ) # the last geno in the group
                     }
    # For the last Group
grp[NGRP]<-  -sum(grp[1:(NGRP-1)])
for(j in NGRP:NGRP){
for (i in first[j]: (last[j]-1) ) { g[i] <- grp[j]+grp.g[i]}
g[last[j]]<-  -sum(grp.g[first[j]:(last[j]-1)] )  # last geno in the last group
}
tau.e<- 1/(sig.e*sig.e)
tau.r<- 1/(sig.r*sig.r)
tau.b<- 1/(sig.b*sig.b)
tau.grp<- 1/(sig.grp*sig.grp)
for(j in 1:NGRP){ tau.g[j] <- 1/(sig.g[j]*sig.g[j]) }
for(j in 1:NGRP){ sig2.g[j] <- (sig.g[j]*sig.g[j])   }
sig2.e <- (sig.e*sig.e)
                    sig2.r<- (sig.r*sig.r)
                    sig2.b<- (sig.b*sig.b)
sig2.grp<- (sig.grp*sig.grp)
 #priors
m ~ dnorm(0.0, 1.0E-6)
for(j in 1:(NGRP-1) ){grp[j]~ dnorm(0, tau.grp)}    # for group effects, except the last grpoup
for(j in 1: NGRP ){
for (i in first[j]: (last[j]-1) ) { grp.g[i] ~ dnorm(0, tau.g[j]) } # geno, except the last geno in the group
                                    }
sig.e ~ dunif(0,100)
for(j in 1:NGRP){     sig.g[j] ~ dunif(0,100) }
sig.r ~ dunif(0,100)
sig.b ~ dunif(0,100)
sig.grp~ dunif(0,100)
   # Prediction of parameters of interest-- means, heritability, SEs
for(j in 1:NGRP) {  h2[j]<- tau.e/(tau.e + tau.g[j]/NR) }
                  # this heritability is on  mean-basis, 8 Jan 2013
    # gain due to 20% selection & 10%, 5% selection K=1.4, 1.755,2.063
for(j in 1:NGRP) { GA20[j] <- 100*1.4/mn/sqrt(tau.g[j]*(1+tau.g[j]/NR/tau.e))  }
```